



Big Data and Reality Mining in Healthcare: Promise and Potential

Hiba Asri¹(✉), Hajar Mousannif², Hassan Al Moatassime¹,
and Jihad Zahir²

¹ OSER Laboratory, Cadi Ayyad University, Marrakesh, Morocco
asri.hiba@gmail.com, hassan.al.moatassime@gmail.com

² LISI Laboratory, Cadi Ayyad University, Marrakesh, Morocco
mousannif@uca.ma, j.zahir@uca.ac.ma

Abstract. Nowadays individuals are creating a huge amount of data; with a cell phone in every pocket, a laptop in every bag and wearable sensors everywhere, the fruits of the information are easy to see but less noticeable is the information itself. This data could be particularly useful in making people's lives healthier and easier, by contributing not only to understand new diseases and therapies but also to predict outcomes at earlier stages and make real-time decisions. In this paper, we explain the potential benefits of big data to healthcare and explore how it improves treatment and empowers patients, providers and researchers. We also describe the capabilities of reality mining in terms of individual health, social network mapping, behavior patterns and treatment monitoring. We illustrate the benefits of reality mining analytics that lead to promote patients' health, enhance medicine, reduce cost and improve healthcare value and quality. Furthermore, we highlight some challenges that big data analytics faces in healthcare.

Keywords: Big data · Reality mining · Healthcare · Predictive analytics · Machine learning

1 Introduction

Individuals are creating torrents of data, far exceeding the market's current ability to create value from it all. A significant catalyst of all this data creation is hyper-specific sensors and smart connected objects (IoT) showing up in everything from clothing (wearable devices) to interactive billboards. Sensors are capturing data at an incredible pace.

In the specific context of healthcare, The volume of worldwide healthcare data is expected to grow up to 25,000 petabytes by 2020 [1]. Some flows have generated 1,000 petabytes of data now and 12 ZBs are expected by 2020 from different sources, such as Electronic Healthcare Record (HER), Electronic Medical Records (EMR), Personal Health Records (PHR), Mobilized Health Records (MHR), and mobile monitors. Moreover, health industries are investing billions of dollars in cloud computing [2]. Thanks to the use of HER and EMR, an estimated number of 50% of hospitals will

integrate analytics solutions, such as Hadoop, hadapt, cloudera, karmasphere, MapR, Neo, Datastax, for health big data [3].

This paper gives insight on the challenges and opportunities related to analyzing larger amounts of health data and creating value from it, the capability of reality mining in predicting outcomes and saving lives, and the Big Data tools needed for analysis and processing. Throughout this paper, we will show how health big data can be leveraged to detect diseases more quickly, make the right decisions and make people's life safer and easier.

The remainder of this paper will be organized as follows:

- Section 2 present the context of reality mining in healthcare system.
- In Sect. 3, we discuss capabilities of reality mining in terms of individual health, social network mapping, behavior patterns and treatment monitoring.
- Sections 4 highlights benefits of reality mining to patients, researchers and providers.
- Section 5 numbers advantages and challenges of big data analytics in the healthcare industry.
- Conclusions and directions for future work are given in Sect. 6.

2 Reality Mining in Healthcare

Reality Mining is about using big data to study our behavior through mobile phone and wearable sensors [4]. In fact every day, we perform many tasks that are almost routines. Cataloguing and collecting information about individuals helps to better understand people's habits. Using machine learning methods and statistical analytics, reality mining can now give a general picture of our individual and collective lives [7].

New Mobile phones are equipped with different kinds of sensors (e.g. motion sensors, location sensors, and haptic sensors). Every time a person uses his/her cell-phone, some information is collected. The Chief Technology Officer of EMC corporation, a big storage company, estimates that the amount of personal sensor information storage will balance from 10% to 90% in the next decade [5]. The most powerful way to know about a person's behavior is to combine the use of some software, with data from phone and from other sources such as: sleep monitor, microphone, accelerometers, camera, Bluetooth, visited website, emails and location [4].

To get a picture of how reality mining can improve healthcare system, here are some examples:

- Using special sensors in mobiles, such as accelerometers or microphone, some diagnosis data can be extracted. In fact, from the way a person talks in conversations, it is possible to detect variations in mood and eventually detect depression.
- Supervising a mobile's motion sensors can contribute to recognize some changes in gait, and could be an indicator of an early stage of Parkinson's disease.
- Using both healthcare sensor data (sleep quality, pressure, heart rate...) and mobile phone data (age, number of previous miscarriage...) to make an early prediction of miscarriage [6].

Communication logs or surveys recorded by mobiles or computers give just a part of the picture of a person's life and behavior. Biometric sensors can go further to track blood pressure, pulse, skin conductivity, heartbeats, brain, or sleep activity. A sign of depression, for instance, can be revealed just by using motion sensors that monitor changes in the nervous system (brain activity) [1].

3 Reality Mining Capabilities

3.1 Individual Health

Currently, the most important source of reality mining is mobile phones. Every time we use our smart phone, we create information. Mobile phone and new technologies are now recording everything about the physical activity of the person. While these data threaten the individual privacy, they also offer a great potential to both communities and individuals.

Authors in [7, 9] assert that the autonomic nervous brain system is responsible of the change of our activity levels that can be measured using motion sensors or by audio; it has been successfully used to screen depression from speech analysis software in mobile phone.

Authors in [10] assert that mobile phones can be used to scale time-coupling movement and speech of the person, which is an indication of a high probability of problems in language development.

Unaware mimicry between people (e.g., posture changes) can be captured using sensors. It is considered as trustworthy indicators of empathy and trusts; and manipulated to strongly enhance compliance. This unconscious mimicry is highly mediated by mirror neurons [11].

Authors in [12] show that several sensors can also detect and measure fluidity and consistency of person's speech or movement. These brain function measurements remain good predictors of human behaviors. Hence, this strong relationship helps for diagnosis of both neurology and psychiatry.

3.2 Social Network Mapping

Besides data from individual health, mobile phones have the capability to capture information about social relationship and social networks. One of the most relevant applications of reality mining is the automatic mapping social network [4]. Through mobile phone we can detect user's location, user's calls, user's SMS, who else is nearby and how is moving thanks to accelerometers integrated in new cell phone.

Authors in [5] describe three type of self-reported: self-reported reciprocal friends when both persons report to other as friend, self-reported non-reciprocal friends when one of both reports to other as a friend and self-reported reciprocal non-friend when no one reports to other as a friend. This information has been shown to be very useful for identifying several important patterns.

Another study use pregnant woman's mobile phone health data like user's activity, user's sleep quality, user's location, user's age, user's Body Mass Index (BMI)among

others, considered as risk factors of miscarriage, in order to make an early prediction of miscarriage and react as earlier as possible to prevent it. Pregnant woman can track her state of pregnancy through a mobile phone application that authors developed [22].

Another good example of network mapping is the computer game named DiabetNet. It is a game intended for young diabetics to help them keep track of their food quality, activity level and blood sugar rates [13].

3.3 Behaviour Patterns and Public Health Service

Behavior pattern involves how a person lives, works, eats etc., not a place, age or residence. Reality mining has the potential to identify behaviors with the use of classification methods that are very useful to predict health patterns [15]. Understanding and combining different behavior patterns of different populations is critical since every subpopulation has its own attitudes and profiles about their health choices.

Google Flu Trends represents a good example to model health of a large population. Just by tracking terms typed in World Wide Web (WWW) searches and identifying the frequency for words related to influenza as illnesses, an influenza outbreak is detected indirectly. In the U.S., Google searches prove an intense correlation between those frequencies and the incidence of estimated influenza based on cases detected by Centers of Disease Control and Prevention (CDC).

Also, with GPS and other technologies, it is easily to track the activities and movements of the person. Location's logs present a valuable source to public health in case of infectious diseases such as tuberculosis, anthrax, and Legionnaires disease. In fact, location logs can help in identifying the source of infections and government may react for preventing further transmission.

3.4 Treatment Monitoring

Once a patient takes his treatment, which is pharmaceutical, behavioral, or other, doctors and clinicians have to monitor their patient's response to treatment. Reality mining information used for diagnosis can be also relevant data for monitoring patient response to treatment and for comparing. Characteristics such as behavior, activity and mobility could be collected in real-time and they could be useful for clinicians to change or adjust treatment depending on patient's response, and in some cases it could be a more effective care with a lower cost.

A concrete example of this is Parkinson's patients. Real-time data are gathered through wearable accelerometers that integrate machine learning algorithms to classify movements' states of Parkinson's patients and to get the development of those movements.

Two significant studies exist in literature to classify dyskinesia, hypokinesia and bradykinesia (slow movements) for seven patients [16]. Data are collected using different sources: wearable accelerometers, videos and clinical observations. Results of studies show that bradykinesia and hypokinesia are the two main classes identified with a high accuracy. Another classification is made to classify patients who feel off or about having dyskinesia.

4 Reality Mining Benefits to Health Care

By combining big data and reality mining, we rang from single to large hospital network. The main benefits can be summarized into detecting diseases at earlier stages, detecting healthcare abuse and fraud faster, and reducing costs. In fact, Big data market also contributes up to 7% of the global GDP and reduces 8% of healthcare costs. Big data analytics improve health care insights in many aspects:

4.1 Benefits to Patients

Big data can help patients make the right decision in a timely manner. From patient data, analytics can be applied to identify individuals that need “proactive care” or need change in their lifestyle to avoid health condition degradation. A concrete example of this is the Virginia health system Carillion Clinic project [17], which uses predictive models for early interventions.

Patients are also more open to giving away part of their privacy if this could save their lives or other people’s lives. “If having more information about me enables me to live longer and be healthier”, said Marc Rotenberg of the electronic Privacy Information Center,” then I don’t think I’m going to object. If having more information about me means my insurance rates go up, or that I’m excluded from coverage, or other things that might adversely affect me, then I may have a good reason to be concerned” [18].

4.2 Benefits to Researchers and Developers (R & D)

Collecting different data from different sources can help improving research about new diseases and therapies [21]. R & D contribute to new algorithms and tools, such as the algorithms by Google, Facebook, and Twitter that define what we find about our health system. Google, for instance, has applied algorithms of data mining and machine learning to detect influenza epidemics through search queries [19, 20]. R & D can also enhance predictive models to produce more devices and treatment for the market.

4.3 Benefits to Healthcare Providers

Providers may recognize high risk population and act appropriately (i.e. propose preventive acts). Therefore, they can enhance patient experience. Moreover, approximately 54% of US hospitals are members in local or regional Health-Information Exchanges (HIEs) or try to be in the future. These developments give the power to access a large array of information. For example, the HIE in Indiana connects currently 80 hospitals and possess information of more than ten million patients [1] (Fig. 1).

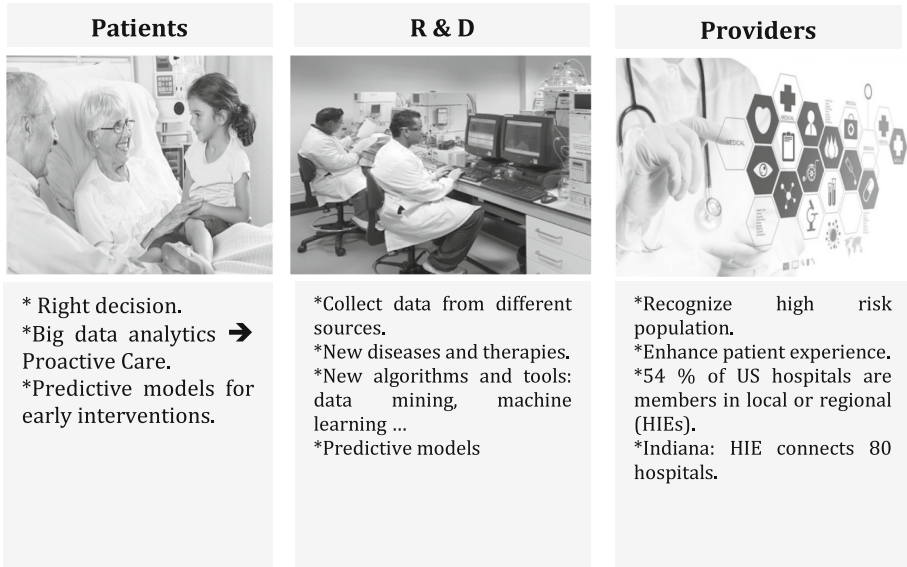


Fig. 1. Summary of reality mining benefits to healthcare.

5 Limitations and Challenges of Data Analytics in Healthcare Industry

Although big data analytics enhance the healthcare industry, there are some limitations to the use of big data in this area:

1. The source of data from organizations (hospital, pharmacies, companies, medical centers...) is in different formats. These organizations have data in different systems and settings. To use this huge amount of data, those organizations must have a common data warehouse in order to get homogeneous information and be able to manage it. However, having such systems requires huge costs.
2. Quality of data is a serious limitation. Data collected are, in some cases, unstructured, dirty, and non-standardized. So, the industry has to apply additional effort to transform information into usable and meaningful data.
3. A big investment is required for companies to acquire staff (data scientists), resources and also to buy data analytics technologies. In addition, companies must convince medical organizations about using big data analytics.
4. Using data mining and big data analytics requires a high level of expertise and knowledge. It is a costly affair for companies to hire such persons.
5. Due to serious constraints regarding the quality of collected data, variations and errors in the results are not excluded.
6. Data security is in big concern and researchers paid more attention on how we can secure all data generated and transmitted. Security problems include personal privacy protection, financial transactions, intellectual property protection and data protection. In some developing and developing countries, they propose laws related

to data protection to enhance the security. So, researchers are asked to carefully consider where they store and analyze data to not be against the regulations.

6 Conclusion and Future Work

Big data is being actively used in healthcare industry to change the way that decisions are made; and including predictive analytics tools, have the potential to change healthcare system from reporting to predicting results at earlier stages. Also reality mining becomes more common in medicine research project due to the increasing sophistication of mobile phones and healthcare wearable sensors. Many mobile phones and sensors are collecting a huge number of information about their user and this will only increase. The use of both Big data and reality mining in healthcare industry has the capability to provide new opportunities with respect to patients, treatment monitoring, healthcare service and diagnosis.

In this survey paper, we discussed capabilities of reality mining in healthcare system including individual health, social network mapping, behavior patterns and public health service, and treatment monitoring; and how patient, providers, researchers and developers benefit from reality mining to enhance medicine. We highlight as well several challenges in Sect. 5 that must be addressed in future works. Adoption of big data and reality mining in healthcare raises many security and patient privacy concerns that need to be addressed.

References

1. Kim, S.G.: Big data in healthcare hype and hope TT - 의료 분야에 있어서 대용량 데이터의 활용. 한국 Cad/Cam 학회 학술발표회 논문집, vol. 2013, pp. 122–125 (2013)
2. Koumpouros, Y.: Healthcare administration. IGI Global (2015)
3. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* **2**(1) (2014). Article number. 3. <https://doi.org/10.1186/2047-2501-2-3>
4. Eagle, N., Pentland, A.S.: Reality mining: sensing complex social systems. *Pers. Ubiquit. Comput.* **10**(4), 255–268 (2006). <https://doi.org/10.1007/s00779-005-0046-3>
5. Eagle, N., Pentland, A.S., Lazer, D.: Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci.* **106**(36), 15274–15278 (2009)
6. Asri, H., Mousannif, H., Al Moatassime, H.: Real-time miscarriage prediction with SPARK. *Procedia Comput. Sci.* **113**, 423–428 (2017)
7. Asri, H., Mousannif, H., Al Moatassime, H., Noel, T.: Big data in healthcare: challenges and opportunities. In: 2015 International Conference on Cloud Technologies and Applications (CloudTech), pp. 1–7. IEEE, June 2015
8. Stoltzman, W.T.: Toward a social signaling framework: activity and emphasis in speech. Doctoral dissertation, Massachusetts Institute of Technology (2006)
9. France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., Wilkes, M.: Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Trans. Biomed. Eng.* **47**(7), 829–837 (2000)
10. Cowley, S.J.: Naturalizing language: human appraisal and (quasi) technology. *AI Soc.* **28**(4), 443–453 (2013). <https://doi.org/10.1007/s00146-013-0445-3>

11. Stel, M., Hess, U., Fischer, A.: The role of mimicry in understanding the emotions of others. In: *Emotional Mimicry in Social Context*, pp. 27–43. Cambridge University Press (2016)
12. Crispiani, P., Palmieri, E.: Improving the fluidity of whole word reading with a dynamic co-ordinated movement approach. *Asia Pac. J. Dev. Differ.* **2**(2), 158–183 (2015)
13. Asri, H., Mousannif, H., Al Moatassime, H.: Reality mining and predictive analytics for building smart applications. *J. Big Data* **6**(1) (2019). Article number. 66. <https://doi.org/10.1186/s40537-019-0227-y>
14. Pouw, I.H.: You are what you eat: serious gaming for type 1 diabetic persons, Master's thesis, University of Twente (2015)
15. Asri, H., Mousannif, H., Al Moatassime, H., Noel, T.: Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Comput. Sci.* **83**, 1064–1069 (2016)
16. Klapper, D.A.: Use of wearable ambulatory monitor in the classification of movement states in Parkinson's disease. Doctoral dissertation, Massachusetts Institute of Technology (2003)
17. IBM News room: IBM predictive analytics to detect patients at risk for heart failure - United States, 19 February 2014
18. Bollier, D.: The Promise and Peril of Big Data, pp. 1–66 (2010)
19. Ghani, K.R., Zheng, K., Wei, J.T., Friedman, C.P.: Harnessing big data for health care and research: are urologists ready? *Eur. Urol.* **66**(6), 975–977 (2014)
20. Lazer, D., Kennedy, R., King, G., Vespignani, A.: The parable of Google Flu: traps in big data analysis. *Science* **343**, 1203–1205 (2014)
21. Asri, H., Mousannif, H., Al Moatassime, H.: Big data analytics in healthcare: case study-miscarriage prediction. *Int. J. Distrib. Syst. Technol. (IJ DST)* **10**(4), 45–58 (2019)
22. Asri, H., Mousannif, H., Al Moatassime, H.: Comprehensive miscarriage dataset for an early miscarriage prediction. *Data Brief* **19**, 240–243 (2018)